Who can be fooled? Modeling perceptions of gullibility from facial appearance

Jaeger, Bastian; Meral, Erdem

Publication date:
2020

Document Version
Early version, also known as pre-print

Who Can Be Fooled? Modeling Perceptions of Gullibility from Facial Appearance

Bastian Jaeger and Erdem O. Meral
Tilburg University

Draft version: 11 April 2020
Word count: 7,603
This paper is currently undergoing peer review. Comment are welcome

Author Note

Bastian Jaeger and Erdem O. Meral, Department of Social Psychology, Tilburg University, The Netherlands.

Both authors contributed equally to this manuscript.

Correspondence concerning this article should be addressed to Bastian Jaeger, Department of Social Psychology, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. E-mail: bxjaeger@gmail.com.

Abstract

In many situations, ranging from cooperative exchange to fraud, people are faced with the challenge to judge how trusting or naïve (i.e., gullible) others are. In three studies, using both theory-driven and data-driven methods, we examine how people form gullibility judgments based on a person's facial appearance. People have a shared representation of what a gullible person looks like. Gullibility impressions are positively related to trustworthiness impressions, but negatively related to dominance impressions (Study 1, $n = 254$). Examining the influence of a wide range of facial characteristics, we find that gullibility impressions are based on cues that have been linked to low levels of perceived threat, such as babyfacedness (Study 2, $n = 403$) and smiles (Study 3, $n = 209$). Together, these findings show that people form gullibility judgments based on facial cues that are seen as indicators of relative harmlessness (i.e., positive intentions and low capabilities).

Keywords: social perception; faces; gullibility; reverse correlation

Who Can Be Fooled? Modeling Perceptions of Gullibility from Facial Appearance

Every year, large amounts of money are lost due to e-mail scams, pyramid schemes, identity thefts, and other types of fraudulent behaviors (Witt, 2018). Aside from the financial costs, fraud victims report experiencing emotional distress including feelings of anger and stress, physical and mental health problems, and issues with their close relationships (Button, Lewis, & Tapley, 2014; Modic & Anderson, 2015). While fraud is common, some groups of individuals, such as the elderly, are targeted more frequently (Burnes et al., 2017; Cohen, 2006). In fact, people differ in how gullible they are (Teunisse, Case, Fitness, & Sweller, 2019). That is, some people may be more at risk because their general tendency to trust and believe others makes them more likely to fall for a scam. How do perpetrators identify gullible targets? Previous research has shown that people spontaneously infer a wide variety of personality traits from a person's facial appearance (Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015). Building on these insights, we investigate how people form impressions of gullibility from facial appearance.

Gullibility Impressions

Personality impressions from faces are formed spontaneously, quickly, and efficiently (Todorov et al., 2015). While people infer a wide variety of traits from a person's facial appearance, three core dimensions capture most variance in face judgments: trustworthiness, dominance, and attractiveness (Oosterhof & Todorov, 2008; Sutherland et al., 2013; Sutherland, Rhodes, Burton, & Young, 2019). These dimensions are thought to reflect the functional significance of trait impressions, with trustworthiness perceptions reflecting an evaluation of a person's intentions, dominance perceptions reflecting the evaluation of a person's ability to implement their intentions, and attractiveness perceptions reflecting the evaluation of a person's mate value (Oosterhof & Todorov, 2008; Sutherland et al., 2013). These impressions can have important consequences (Olivola, Funk, & Todorov, 2014). For instance, trustworthiness impressions influence criminal sentencing decisions (Wilson & Rule, 2015), personnel selection (Gomulya, Wong, Ormiston, & Boeker, 2017), and interpersonal trust (Jaeger, Evans, Stel, & van Beest, 2019). In short, people spontaneously infer personality traits from facial appearance and these inferences guide a wide range of social decisions.

While a wealth of studies has examined what makes a person look trustworthy, little is known about what makes a person look trusting or gullible. Do people think that individuals with a certain facial appearance are particularly gullible? Given the lack of previous studies on

gullibility and facial stereotypes, we adopt a relatively broad conceptualization of gullibility here. We define a gullible individuals as someone who is very trusting (rather than distrusting), who is very naïve (rather than skeptical), and who easily believes others (rather than is easily suspicious of others; Krueger, Vogrincic-Haselbacher, & Evans, 2019; Teunisse et al., 2019; Yamagishi, Kikuchi, & Kosugi, 1999). The goals of the current investigation were threefold. First, we examine whether people share stereotypes about the typical facial appearance of a gullible person. If people have shared representation of a gullible face, then they should show at least some level of consensus in their judgments (Hehman, Sutherland, Flake, & Slepian, 2017).

Second, we examine relationships between gullibility impressions and impressions of trustworthiness and dominance. Prior work suggests that gullible people are seen as well-intentioned and approachable, but also as relatively incompetent (Evans & van de Calseyde, 2018). This view predicts that that gullible-looking people should score high on perceived trustworthiness (a proxy for an individual's perceived intentions), but low on perceived dominance (a proxy for an individual's ability to implement their intentions; Todorov, Said, Engell, & Oosterhof, 2008).

Finally, we examine which facial characteristics make a person look gullible. People rely on various features when forming trait impressions from faces (Todorov et al., 2015). For instance, perceptions of trustworthiness are driven by morphological characteristics such as facial width-to-height ratio (Ormiston, Wong, & Haselhuhn, 2017), statistical characteristics such as the averageness or sex-typicality (Dotsch & Todorov, 2012; Sofer, Dotsch, Wigboldus, & Todorov, 2015), and resemblances between facial features and emotional expressions (Said, Sebe, & Todorov, 2009). Here, we test how a wide range of facial cues—which have been linked to trait impressions in previous studies—influence perceptions of gullibility.

The Current Studies

In three preregistered studies, we model perceptions of gullibility using different sets of face stimuli and statistical techniques. In Study 1 (n = 254), we rely on reverse correlation—a data-driven technique—to create images of prototypically gullible-looking faces (Dotsch, Wigboldus, Langner, & Van Knippenberg, 2008). We also recruit a separate sample of participants (n = 40) to test how gullibility perceptions are related to perceptions of other core dimensions of social perception.

In Studies 2 and 3, we measure perceptions of gullibility for two sets of real, rather than computer-generated faces. In both studies, we test whether there is a shared gullibility stereotype by estimating the level of consensus in participants' judgments. In Study 2 (n = 403) we measure gullibility perceptions of 183 emotionally neutral faces. We test how a wide range of facial characteristics that are commonly studied in the social perception literature influence gullibility perceptions.

In Study 3 (n = 209) we measure gullibility perceptions of 342 faces. Going beyond emotionally neutral faces, we examine whether smiles, which are typically seen as a sign of positive intentions (Martin, Rychlowska, Wood, & Niedenthal, 2017; Mehu, Grammer, & Dunbar, 2007), increase perceived gullibility. Moreover, previous work on the demographic characteristics of fraud victims suggest that young and old (vs. middle-aged) people and women are more likely to be targets of fraudulent behavior (Jagatic, Johnson, Jakobsson, & Menczer, 2007; Sheng, Holbrook, Kumaraguru, Cranor, & Downs, 2010). We therefore test whether this pattern is also reflected in perceptions of gullibility.

All data, analysis scripts, materials, and preregistration documents are available at the Open Science Framework (https://osf.io/tx6p3/). We report how our sample sizes were determined and all data exclusions and measures for each study.

Study 1

The objectives of Study 1 were to (a) obtain a visual image of what people think a gullible face looks like and (b) test how perceptions of gullibility are related to perceptions of other core trait dimensions (Oosterhof & Todorov, 2008; Sutherland et al., 2013). To this end, participants completed a reverse-correlation image classification task (Dotsch & Todorov, 2012). In this task, individuals are presented with two facial photographs, both depicting the same face with average features. One image is superimposed with a random noise pattern and the other image is superimposed with the reverse of the random noise pattern. This creates slight and completely random differences in facial appearance. Across a large number of trials in which noise patterns are varied, participants are asked to choose the face that looks more gullible. Afterwards, all noise patterns of the selected images are averaged. That is, the noise patterns that were judged to produce a more gullible-looking facial appearance are combined and superimposed on the base image, producing an approximation of a prototypically gullible-looking face. The noise patterns that were not selected are also combined to produce an anti-

gullible face. This approach has been successfully used to visualize a variety of facial stereotypes for traits such as trustworthiness, dominance, warmth, and competence, (Dotsch & Todorov, 2012; Imhoff, Woelki, Hanke, & Dotsch, 2013; Sofer et al., 2015).

In the second phase (which was preregistered), an independent sample of participants rated the prototype images. We attempted to validate the results of the reverse correlation task by testing whether the gullibility prototype is indeed rated as more gullible than the anti-gullibility prototype. We also obtained ratings of trustworthiness, dominance, and attractiveness to test how gullible facial features correlate with perceptions on core dimensions of social perception. We predicted that the gullibility prototype would be seen as more trustworthy, but less dominant than the anti-gullibility prototype. We did not have strong expectations regarding differences in attractiveness. However, as attractiveness ratings are correlated with trustworthiness ratings (and other evaluations of a person's intentions) in computer-generated faces (Oosterhof & Todorov, 2008), we predicted higher attractiveness for the gullibility prototype,

Methods

Image construction. In the image construction phase, we recruited 255 students who completed the study for partial course credit. One participant (0.39%) was excluded due to missing data, leaving a final sample of 254 participants (M_{age} = 19.62, SD_{age} = 1.91, 89.76% female). Participants completed a two images forced choice variant of the reverse correlation task (Dotsch & Todorov, 2012; Dotsch et al., 2008). On every trial, two images of the same base face (a morphed image of 50 male faces from the Karolinska Directed Emotional Faces database Lundqvist, Flykt, & Öhman, 1998) were displayed. One image was superimposed with a random noise pattern, whereas the other image was superimposed with the negative of the random noise pattern. The position of the two images and the order in which the different noise patterns were displayed were randomized.

Across a total of 300 trials, participants selected the face that most resembled a gullible person. Participants were told that a person who scores high on gullibility "*is very trusting, naïve, and easily believes others*" and that a person who scores low on gullibility "*is very distrusting, skeptical, and easily suspicious of others.*"

We created a group-level classification image (CI) by averaging the noise patterns on the images that participants had chosen as more gullible. We also created a group-level anti-CI by

averaging the noise patterns of the images that participants did not choose. In this case, the anti-CI is supposed to represent a person who is not gullible.

Image rating. In the image rating phase, we recruited 40 American Mechanical Turk workers (37.50% female, M_{age} = 33.90, SD_{age} = 7.54) to rate the gullibility and anti-gullibility prototypes on four trait dimension: gullibility, trustworthiness, dominance, and attractiveness. All participants first indicated their gullibility ratings and then their ratings for the other three dimensions (in a randomized order). We displayed one image at a time and the position of the two face prototypes was counterbalanced. Participants indicated their responses on a scale that ranged from 1 (*not at all* [trait]) to 9 (*extremely* [trait]).

Results

The face prototypes resulting from the reverse correlation task are displayed in Figure 1. The gullibility prototype is shown on the left and the anti-gullibility prototype is shown on the right. A paired-samples *t*-test showed that the gullibility prototype was indeed seen as more gullible (M = 6.30, SD = 1.88) than the anti-gullibility prototype (M = 4.55, SD = 2.28), $t(39)$ = 3.21, p = .003, Hedge's g = 0.50.

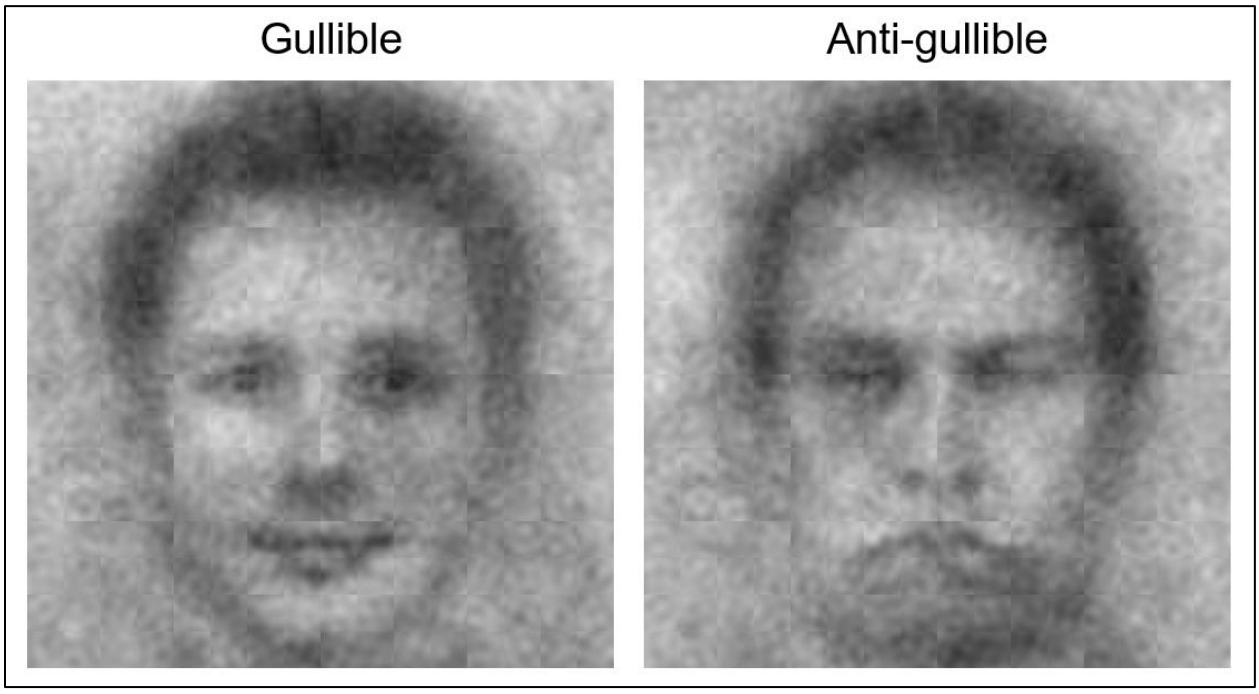

Figure 1. Resulting classification images for the gullibility prototype (left) and the anti-gullibility prototype (right)

Next, we examined how the two prototypes were rated on trustworthiness, dominance, and attractiveness. In line with our predictions, the gullible face was seen as more trustworthy (M = 7.00, SD = 1.22) than the anti-gullible face (M = 3.98, SD = 2.21), $t(39) = 7.18$, $p < .001$, Hedge's $g = 1.12$. The gullible face was also seen as less dominant (M = 4.65, SD = 1.99) than the anti-gullible face (M = 6.63, SD = 1.63), $t(39) = 4.53$, $p < .001$, Hedge's $g = 0.71$. Finally, the gullible face was seen as more attractive (M = 6.55, SD = 1.47) than the anti-gullible face (M = 3.68, SD = 2.13), $t(39) = 7.83$, $p < .001$, Hedge's $g = 1.44$.

Discussion

Study 1 provided initial insights into what people think a gullible person looks like. A visual inspection of the prototypes suggests that a gullible-looking person has baby-faced features, with a smiling mouth, and widely opened eyes, whereas the anti-gullibility face looks older, with downturned corners of the mouth, lowered eyebrows, and partly closed eyes. Ratings by an independent sample of participants confirmed that the gullibility prototype was indeed seen as more gullible than the anti-gullibility prototype, suggesting that people somewhat agree on which facial features make a person look gullible. Moreover, the gullible face was also seen as more trustworthy and less dominant than the anti-gullible face. These findings are in line with the notion that gullible people are seen as well-intentioned and harmless individuals (Evans & van de Calseyde, 2018).

Study 2

In Study 2, we extended findings of Study 1 in three important ways. First, rather than using computer-generated images, we used naturalistic face images. This allowed us to test how naturally occurring variations in facial features influence gullibility perceptions. Second, we examined how a wide range of facial characteristics influence gullibility perceptions. Specifically, we focus on four classes of characteristics, which have been studied extensively in the social perception literature: demographic characteristics (e.g., gender and age), emotion resemblances (e.g., resemblance to an angry facial expressions), psychological characteristics (e.g., attractiveness or femininity), and morphological characteristics (e.g., facial width-to-height ratio). Finally, results of Study 1 suggest that perceptions of gullibility are related to perceptions of trustworthiness and dominance, which represent core dimensions on which faces are evaluated (Oosterhof & Todorov, 2008). This raises the question of whether facial stereotypes of gullibility can be dissociated from facial stereotypes of trustworthiness and dominance. We examined this

question by testing whether facial characteristics differently predict judgments of gullibility, trustworthiness, and dominance. To answer these questions, we leveraged the extensive norming data of the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015). The database contains photographs of a large and diverse set of targets, and measurements of a wide range of facial characteristics.

Methods

This study was preregistered and we explicitly mention if a decision or analysis was not specified a priori.

Participants. Simulation results show that data from 40 independent raters yield stable average ratings even when consensus is relatively low (Hehman, Xie, Ofosu, & Nespoli, 2018). We therefore aimed to recruit at least 120 participants (i.e., 40 raters per image set), with the final sample size being determined by how many participants completed the study within two weeks. In total, we recruited 512 first-year psychology students from a Dutch university who completed the study in return for partial course credit. In line with our preregistration, data from 102 participants (19.92%) who indicated poor or basic English proficiency, from 3 participants (0.01%) who always indicated the same rating across all trials, and from 4 participants (0.01%) whose response time was shorter than 100 milliseconds on at least 10% of all trials were excluded from analyses, leaving a final sample of 403 participants (M_{age} = 20.00, SD_{age} = 2.12; 80.15% female).

Stimuli. We selected all 183 images of Caucasian adults (93 male and 90 female) from the Chicago Face Database (Ma et al., 2015). All targets displayed a neutral facial expression and were photographed from a fixed distance against a uniform background. The estimated age of targets ranged from 17 to 50 (M_{age} = 27.87, SD_{age} = 5.84). We created three sets of stimuli. Each stimulus set contained 61 images (30 female).

Procedure. Participants were randomly assigned to one stimulus set. They were asked to rate how gullible they think the person in the photo is on a scale from 1 (*not at all*) to 7 (*extremely*). We used the same gullibility description as in Study 1. We created average gullibility ratings per image. These averages were based on ratings by a minimum of 129 participants.

Facial characteristics. For each target, the Chicago Face Database contains data on an extensive set of facial characteristics. Here, we focused on four sets of facial characteristics,:

demographic variables (gender and age), resemblances to emotional expressions (the extent to which facial features resemble an expression of fear, anger, disgust, happiness, sadness, and surprise), psychological traits (perceived attractiveness, trustworthiness, dominance, femininity, babyfacedness, prototypicality, and unusualness), and morphological traits (luminance, fWHR, lip fullness, eye size, and nose shape). Data on gender and race were provided by the photographed individuals. Data on emotional resemblances and psychological traits represent subjective ratings provided by participants, with each characteristic representing the mean rating from an average of 44 independent raters (*Min* = 21, *Max* = 131). Morphological features were measured in Adobe Photoshop. A more detailed description of the variables and how they were measured is provided by Ma and colleagues (Ma et al., 2015).

Results

Consensus. First, we computed intraclass correlation coefficients (*ICC*s) to estimate consensus in gullibility ratings across participants (Shrout & Fleiss, 1979). Participants showed significant consensus in their judgments of set 1, $ICC(2, 1) = .062$, $p < .001$, 95% CI [.043, .092], set 2, $ICC(2, 1) = .095$, $p < .001$, 95% CI [.068, .138], and set 3, $ICC(2, 1) = .077$, $p < .001$, 95% CI [.055, .113].

Gullibility perceptions. Next, we examined which facial characteristics drive perceptions of gullibility. We estimated separate regression models in which we regressed gullibility ratings on demographic variables (e.g., gender and age), emotion resemblances (e.g., resemblance to a facial expression of anger or happiness), psychological traits (e.g., perceived attractiveness or), and morphological traits (e.g., fWHR or eye size). All variables (including gullibility ratings but excluding age) were *z*-standardized prior to analysis. Within each model, we applied Bonferroni-Holm correction to account for multiple comparisons. This step was not specified in our preregistration, but represents a more conservative test given the relatively large number of predictors. For the sake of brevity, we only report detailed statistics for variables that significantly predicted perceptions of gullibility. Regression estimates for all predictors can be found in Table 1 (column 1).

Demographic variables explained 25% of the variance in gullibility perceptions, $F(2, 180) = 30.53$, $p < .001$, adjusted $R^2 = .245$. Women were seen as more gullible than men, $\beta =$

0.355, *SE* = 0.129, *t*(180) = 2.76, *p* = .006, and younger targets were seen as more gullible than older targets, β = 0.082, *SE* = 0.011, *t*(180) = 7.43, *p* < .001.[1]

Resemblance to emotional expressions explained 51% of the variance in gullibility perceptions, *F*(6, 176) = 32.96, *p* < .001, adjusted R^2 = .513. Faces that resembled an expression of anger were seen as less gullible, β = -0.759, *SE* = 0.110, *t*(176) = 6.89, *p* < .001. There were no significant effects for resemblance to expressions of fear, disgust, happiness, sadness, or surprise.

Perceptions of psychological traits explained 63% of variance in gullibility perceptions, *F*(7, 175) = 44.61, *p* < .001, adjusted R^2 = .627. Faces that were perceived as trustworthy, β = 0.325, *SE* = 0.065, *t*(175) = 5.04, *p* < .001, and babyfaced, β = 0.338, *SE* = 0.053, *t*(175) = 6.42, *p* < .001, were also seen as gullible. Perceptions of dominance were negatively associated with perceptions of gullibility, β = -0.484, *SE* = 0.062, *t*(175) = 7.87, *p* < .001. There were no significant effects for perceptions of attractiveness, femininity, race-typicality, and unusualness.

Morphological characteristics only explained 8% of variance in gullibility perceptions, *F*(5, 177) = 4.25, *p* = .001, adjusted R^2 = .082. Faces with large, opened eyes were seen as more gullible, β = 0.296, *SE* = 0.076, *t*(177) = 3.91, *p* = .001. There were no significant effects for luminance, fWHR, lip fullness, and nose shape.

Differences with trustworthiness and dominance perceptions. Replicating findings of Study 1, the present results suggest that perceptions of gullibility are positively related to perceptions of trustworthiness and negatively related to perceptions of dominance. This raises the question of how much these stereotypes overlap. To address this question, we examined to what extent perceptions of gullibility, trustworthiness and dominance are based on the same facial features by testing whether specific facial features differently predicted ratings on the three dimensions. First, we estimated separate models in which we regressed trait ratings on a dummy variable indicating which trait was judged (coded 0 for gullibility and 1 for trustworthiness, for example), the relevant set of facial features (e.g., all demographic variables), and an interaction between the trait dummy and a specific facial feature (e.g., gender). These models also included a random intercept per target, which accounts for differences in average ratings across targets irrespective of which trait was judged. This allowed us to test whether a given facial feature

[1] We also estimated a separate model in which we included a quadratic term for the effect of age, but this did not yield a significant effect, β = 0.002, *SE* = 0.001, *t*(179) = 1.56, *p* = .12.

differently predicted perceptions of gullibility and trustworthiness. We again applied Bonferroni-Holm correction to account for multiple comparisons within each model. We followed the same procedure to examine differences between gullibility and dominance perceptions.[2]

The estimates for each interaction effect and the trait dummy variable (trustworthiness vs. gullibility or dominance vs. gullibility) are displayed in Table 1. These analyses showed that many features (9 out of the 18 examined here) differently predicted perceptions of gullibility and trustworthiness (see Table 1, column 2). Gender and age; resemblances to happy and sad facial expressions; and perceptions of attractiveness, femininity, babyfacedness, prototypicality, and unusualness were differently related to ratings of gullibility and trustworthiness. To illustrate, older targets were seen as less gullible and less trustworthy, but age was stronger determinant of gullibility perceptions than trustworthiness perceptions. In other words, when forming impressions based on facial features, people rely more on age when judging gullibility than when judging trustworthiness.

A similar pattern was observed when comparing perceptions of gullibility and dominance, with 9 out of 18 facial features differently predicting ratings of the two traits. Gender and age; resemblances to angry, disgusted, and happy facial expressions; perceptions of femininity and bayfacedness; and luminance and eye size were differently related to ratings of gullibility and dominance. In sum, these results show that different facial features drive perceptions of gullibility, trustworthiness, and dominance, which suggests that they represent at least somewhat dissociable facial stereotypes.

[2] In our preregistration, we only specified to examine differences between gullibility and trustworthiness because a visual inspection of the gullibility prototype in Study 1 appeared most similar to trustworthiness prototypes obtained in a previous study (Dotsch & Todorov, 2012). However, given that the gullibility and anti-gullibility prototypes also significantly differed on perceived dominance, which is often treated as the second core dimension in trait impressions from faces (Oosterhof & Todorov, 2008), we decided to also conduct analyses examining differences between gullibility and dominance. At any rate, we did not have any a priori expectations regarding differences in trait perceptions and these results should be treated as exploratory.

Table 1
Determinants of gullibility perceptions and trustworthiness perceptions.

	Gullibility	Trustworthiness -Gullibility	Dominance -Gullibility
Demographics			
Gender	0.355 **	0.407 **	-1.030 ***
Age	-0.082 ***	0.053 ***	0.124 ***
Emotion resemblance			
Fear	0.209	-0.194	-0.203
Anger	-0.759 ***	0.113	1.253 ***
Disgust	0.044	0.023	0.848 ***
Happiness	0.103	0.328 ***	-0.558 ***
Sadness	0.135	-0.287 **	-0.160
Surprise	0.016	0.083	-0.213
Psychological traits			
Attractiveness	-0.133	0.660 ***	0.114
Trustworthiness	0.325 ***	-	-
Gullibility	-	-	-
Dominance	-0.484 ***	-	-
Femininity	-0.055	0.269 **	-0.635 ***
Babyfacedness	0.338 ***	-0.300 ***	-0.912 ***
Prototypicality	-0.089	0.159 *	-0.232
Unusualness	0.028	-0.184 *	0.281
Morphological traits			
Luminance	0.155	-0.064	-0.382 **
fWHR	-0.002	-0.018	0.175
Lip fullness	0.054	-0.091	-0.153
Eye size	0.296 **	0.0001	-0.594 ***
Nose shape	0.068	-0.141	0.132

*** $p < .001$. ** $p < .01$. * $p < .05$.

Discussion

Study 2 provided more insights into which facial features contribute to perceptions of gullibility. We examined the influence of a wide range of characteristics that have previously been shown to form the basis for other trait impressions (Said et al., 2009; Stirrat & Perrett, 2010; Todorov et al., 2008). Specifically, we examined 20 characteristics representing 4 dimensions that are commonly used to describe faces: demographic characteristics, resemblances to emotion expressions, psychological characteristics, and morphological characteristics. Results showed that women and younger individuals (within the age range of 17 to 50) were seen as more gullible. Individuals whose resting facial features resembled an angry facial expression were seen as less gullible. Trustworthy-looking individuals and individuals with more child-like facial features were seen as more gullible, whereas dominant-looking individuals were seen as less gullible. Finally, individuals with large, opened eyes were also seen as more gullible.

Our results also showed that perceptions of gullibility are at least somewhat distinct from perceptions of trustworthiness and dominance, with many facial characteristics differently influencing judgments of the three traits. For instance, resemblance to an angry facial expression was related to gullibility (but not trustworthiness) judgments, whereas resemblance to a happy facial expression was related to trustworthiness (but not gullibility) judgments. In a similar vein, babyfacedness predicted perceptions of gullibility (but not dominance), whereas facial femininity predicted perceptions of dominance (but not gullibility). These results suggest that facial gullibility stereotypes are at least somewhat dissociable from facial trustworthiness and dominance stereotypes.

Study 3

Study 3 extended our previous analyses in three ways. First, Study 1 and Study 2 focused on perceptions of gullibility in neutral faces. However, facial expressions—and smiles in particular—are common in everyday life (Martin et al., 2017), especially in situations in which people might be motivated to evaluate a person's gullibility (e.g., social networking or dating websites). We therefore examined the effect of smiling on gullibility perceptions. Smiles are usually seen as a signal of affiliation and positive intentions (Martin et al., 2017; Mehu et al., 2007). We therefore predicted that smiling individuals would be seen as more gullible than emotionally neutral individuals.

Second, previous studies suggest that young and old (vs. middle-aged) individuals are particularly likely to be targets of fraudulent behavior (Cohen, 2006; Jagatic et al., 2007; Sheng et al., 2010). In Study 2, we found a negative linear effect of age on perceived gullibility, but no quadratic effect. That is, young, but not old individuals were seen as more gullible. However, the oldest individual in the image set was only 50 years old. In Study 3, we therefore examined perceptions of gullibility across a wider age range. We tested whether young adults (ca. 20 to 30 years old) and old adults (ca. 70 to 80 years old) would be seen as more gullible than middle-aged adults (ca. 35 to 45 years old).

Third, we again examined gender differences in perceived gullibility. In Study 2, we found that women were seen as more gullible than men when using a stimulus set of U.S. American individuals. Here, we tested whether this finding replicates in a stimulus set of German individuals who varied more in their age.

Methods

This study was preregistered and we explicitly mention if a decision or analysis was not specified a priori.

Participants. We again aimed to recruit at least 40 independent raters per image set (i.e., a total of 200 participants, Hehman, Xie, Ofosu, & Nespoli, 2018), with the final sample size being determined by how many participants completed the study within two weeks. In total, we recruited 212 first-year psychology students from a Dutch university who completed the study in return for partial course credit. In line with our preregistration, data from one participant (0.47%) who always indicated the same rating across all trials, and from two participants (0.95%) whose response time was shorter than 100 milliseconds on at least 10% of all trials were excluded from analyses, leaving a final sample of 209 participants (M_{age} = 20.19, SD_{age} = 2.39; 72.73% female, 26.79% male, 0.48% other). We also preregistered to exclude participants who indicate poor or basic English proficiency. However, due to a technical error, data on English proficiency was not assessed.

Stimuli. We selected 171 images of Caucasian adults from the FACES Database (Ebner, Riediger, & Lindenberger, 2010). We selected images of 58 young targets (M_{age} = 24.2, SD_{age} = 3.4, 29 female), 56 middle-aged targets (M_{age} = 49.0, SD_{age} = 3.9, 27 female), and 57 old targets (M_{age} = 73.2, SD_{age} = 2.8, 29 female). All targets were photographed from a fixed distance against a uniform background. For each target, we selected an image in which they were

displaying a neutral facial expression and an image in which they were displaying a happy facial expression. Thus, the total number of images was 342 with gender (male vs. female) and age group (young vs. middle-aged vs. old) varying between targets and facial expression (neutral vs. smiling) varying within targets. We created five sets of stimuli. Each stimulus set contained either 63 or 72 images with approximately equal numbers of male and female targets and young, middle-aged, and old targets.

Procedure. Participants were randomly assigned to one stimulus set. They were asked to rate how gullible they think the person in the photo is on a scale from 1 (*not at all*) to 7 (*extremely*). We again used the same gullibility description as in Studies 1 and 2.

Results

Consensus. First, we computed intraclass correlation coefficients (*ICC*s) to estimate consensus in gullibility ratings across participants. Participants showed significant consensus in their judgments of set 1, $ICC(2, 1) = .320$, $p < .001$, 95% CI [.249, .414], set 2, $ICC(2, 1) = .210$, $p < .001$, 95% CI [.158, .282], set 3, $ICC(2, 1) = .327$, $p < .001$, 95% CI [.260, .415], set 4, $ICC(2, 1) = .254$, $p < .001$, 95% CI [.197, .333], and set 5, $ICC(2, 1) = .170$, $p < .001$, 95% CI [.123, .239].

Gullibility perceptions. Next, we estimated a multilevel regression model in which we regressed gullibility ratings on facial expression (neutral vs. smiling), gender (male vs. female), and age (young vs. middle-aged vs. old, with middle-aged as the reference category; see Figure 2). We included random intercepts for participants and targets (to account for differences in average gullibility ratings across participants and targets) and random slopes for all predictors (to account for variance in effects across participants and targets). In line with our predictions, we found that smiling individuals were perceived as more gullible than neutral individuals, $\beta = 1.546$, $SE = 0.096$, $t(252.3) = 16.10$, $p < .001$, and women were seen as more gullible than men, $\beta = 0.218$, $SE = 0.049$, $t(200.8) = 4.47$, $p < .001$. We also found that young individuals were seen as more gullible than middle-aged individuals, $\beta = 0.183$, $SE = 0.060$, $t(162.0) = 3.04$, $p = .003$. However, contrary to our expectations, we found no significant difference in perceived gullibility between older and middle-aged individuals, $\beta = -0.017$, $SE = 0.062$, $t(140.9) = 0.28$, $p = .78$.

We also explored whether there were any interaction effects between facial expression, gender, and age. We again estimated a multilevel regression model with random intercepts per participant and target. This time, we included all two-way and three-way interaction effects

between the three variables. There were significant interaction effects between facial expression and age. To probe these interaction effects, we examined the effect of facial expression for young, middle-aged, and old targets. Smiling (vs. neutral) individuals were seen as more gullible when they were young, β = 1.264, *SE* = 0.100, *t*(197.3) = 12.63, $p < .001$, middle-aged, β = 1.626, *SE* = 0.110, *t*(187.2) = 14.77, $p < .001$, and old, β = 1.753, *SE* = 0.115, *t*(182.9) = 15.23, $p < .001$. However, the effect of a happy facial expression on gullibility perceptions was significantly smaller for young (vs. middle-aged) individuals, β = -0.321, *SE* = 0.123, *t*(156.9) = 2.62, $p = .010$, and significantly larger for old (vs. middle-aged) individuals, β = 0.269, *SE* = 0.124, *t*(157.1) = 2.18, $p = .031$. No other two-way or three-way interactions were significant.

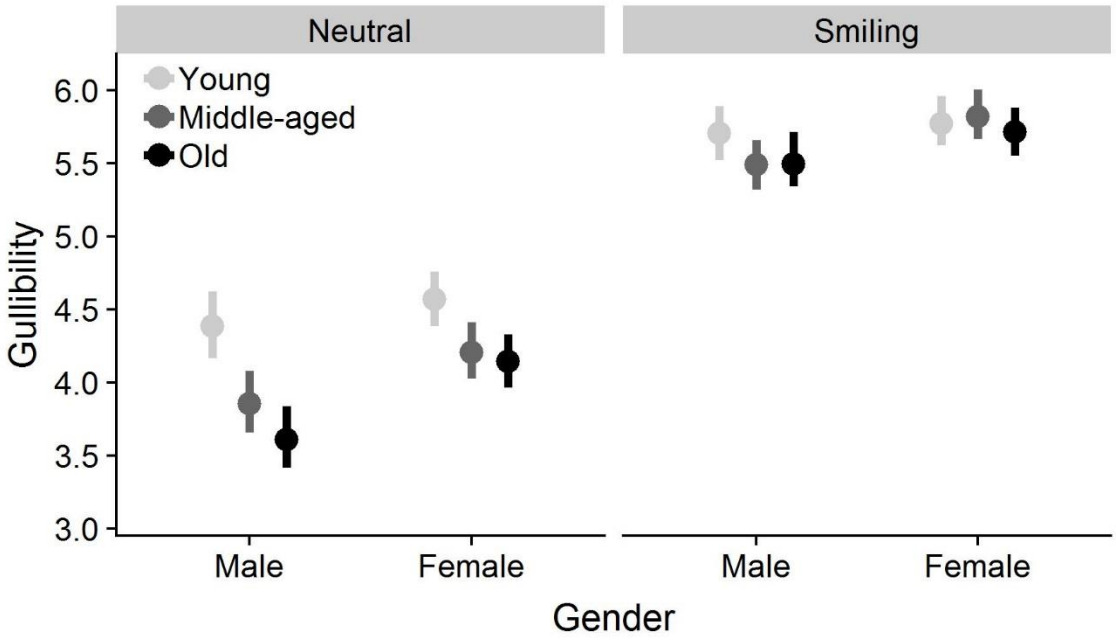

Figure 2. The influence of gender, age, and facial expression on perceptions of gullibility. Dots denote predicted gullibility ratings derived from a multilevel regression model. Error bars denote bootstrapped 95% confidence intervals.

Discussion

In the current study, we examined how gender, age, and facial expression (i.e., smiling vs. neutral) contribute to perceptions of gullibility. We found that women were seen as more gullible than men and that young individuals (ca. 20 to 30 years old) were seen as more gullible than middle-aged individuals (ca. 35 to 45 years old). Contrary to our expectations, we did not

find that old individuals (ca. 70 to 80 years old) were seen as more gullible than middle-aged individuals. We also examined how smiling affects perceptions of gullibility. Results of Studies 1 and 2 showed that gullibility judgments are positively correlated with trustworthiness judgments, suggesting that they reflect, at least partly, an appraisal of another person's intentions. As smiles signal affiliation and positive intentions (Martin et al., 2017; Mehu et al., 2007), we predicted that smiling individuals would be seen as more gullible. Results were in line with this hypothesis. In fact, smiling individuals were always seen as more gullible regardless of their gender or age.

General Discussion

What makes a person look gullible? Across three preregistered studies using both data-driven and theory-driven methods, we examined how people form impressions of gullibility based on facial appearance. We investigated (a) whether people share stereotypes regarding what a gullible person looks like, (b) how gullibility impressions are related to impressions of core dimensions of social perception (i.e., trustworthiness and dominance), and (c) which facial characteristics people rely on to form gullibility impressions.

Across three studies, we consistently find that people at least somewhat agree on what a gullible person looks like. In the reverse correlation task in Study 1, people repeatedly selected which random noise pattern increased perceived gullibility when superimposed on a face. Creating separate average noise patterns across all selected and unselected ones clearly produced faces with different appearances (see Figure 1). This suggests the presence of a shared understanding of what the target category (i.e., gullibility) looks like because the absence of a common stereotype would lead to a random or highly idiosyncratic selection of noise patterns, which, when averaged, would result in two identical faces. The idea that people have a shared representation of a gullible face was also corroborated by analyzing gullibility ratings of naturalistic faces in Studies 2 and 3. Results showed a significant level of consensus in gullibility judgments (as indicated by significant intraclass correlation coefficients). Taken together, these results suggest that people show some level of agreement in what a gullible person looks like.

While people can infer a variety of traits from faces (Oosterhof & Todorov, 2008), not every trait is characterized by a unique facial appearance. In fact, previous studies have shown that faces are primarily judged along three dimensions: trustworthiness (representing judgments of a target's intentions), dominance (representing judgments of a target's abilities), and

attractiveness (representing judgments of a target's mate value; Oosterhof & Todorov, 2008; Sutherland et al., 2013). How were perceptions of gullibility related to these core dimensions of social perception? Gullibility is associated with approachability and positive intentions, but also with low competence and harmlessness (Evans & van de Calseyde, 2018). In line with this view, perceptions of gullibility were positively related to perceptions of trustworthiness, but negatively to perceptions of dominance (Studies 1 and 2).

Zero-order correlations between perceptions of gullibility and trustworthiness (r = .47) and between perceptions of gullibility and dominance (r = -.68) were relatively strong. These results suggest that gullibility perceptions do not neatly fall into one of these two core dimensions of social perception. In fact, we found that many facial features were differently related to perceptions of gullibility, trustworthiness, and dominance. For example, the extent to which neutral faces resembled an expression of anger influenced perceptions of gullibility (but not trustworthiness), whereas resemblance to a happy facial expression influenced perceptions of trustworthiness (but not gullibility). Moreover, women were seen as both more gullible and more trustworthy than men, but the effect of gender was significantly stronger for trustworthiness impressions. Similarly diverging patterns were observed when comparing gullibility and dominance perceptions. Together, these results suggest that perceptions of gullibility, trustworthiness, and dominance represent dissociable facial stereotypes.

Which facial cues do people rely on to form gullibility judgments? Both data-driven (Study 1) and theory-driven (Study 2) approaches yielded converging results: People who are perceived as gullible are young and female with babyfaced features, including large, opened eyes. People whose facial features resemble an expression of anger are perceived as particularly low on gullibility. Going beyond emotionally neutral faces, Study 3 also investigated the influence of smiles on gullibility perceptions. Smiles are typically interpreted as signals of affiliation and positive intentions (Martin et al., 2017; Mehu et al., 2007) and should therefore have a positive influence on perceived gullibility. Indeed, we found that irrespective of their gender or age, smiling individuals were seen as more gullible that emotionally neutral individuals. Taken together, findings from all three studies suggest that when forming impressions of gullibility, people rely on a suite of facial cues that are seen as indicating high levels of approachability and low levels of threat.

Limitations and Future Directions

The present studies showed that people rely on various facial features to form judgments of gullibility. Do people actually rely on these cues, for example, when deciding whom to target for financial scams? Previous work on the demographic characteristics of fraud victims suggest that this might indeed be the case. For instance, women and young people are more likely to fall victim to scams (Jagatic et al., 2007; Sheng et al., 2010). However, the influence of gullibility perceptions probably extends beyond situations in which one party is attempting to exploit another. In many mixed-motive situations—such as the prisoner's dilemma (Camerer, 2003; Rapoport & Chammah, 1970)—mutual gains are only realized if both parties trust the other to cooperate. No matter whether a person wants to establish a cooperative relationship or exploit the other's trust, a necessary requirement for both goals is that the other party cooperates. Additional studies are needed to test whether people with a gullible facial appearance are indeed more likely to be sought out in such situations.

Future studies could also examine the accuracy of gullibility perceptions. It is possible that people agree on what a gullible person looks like and that they rely on their gullibility inferences when making decisions even though these judgments are not correlated with a person's actual gullibility. In fact, many studies have shown that people rely on impressions of trustworthiness, competence, or other personality traits, even though the accuracy of these impressions is very low (Bonnefon, Hopfensitz, & De Neys, 2017; Todorov et al., 2015). More work is needed to test whether gullibility perceptions contain a kernel of truth.

In the present studies, we conceptualized gullibility very broadly. In all three studies, participants indicated their gullibility perceptions by rating to what extent they believe that a target is very trusting, naïve, and easily believes others (high gullibility) or very distrusting, skeptical, and easily suspicious of others (low gullibility). One might argue that this definition conflates gullibility and trust, which have been treated as separate constructs in prior research (Rotter, 1967; Teunisse et al., 2019; Yamagishi et al., 1999). For example, Teunissen and colleagues (2019) argued that trust is a generalized expectation that others are trustworthy (in absence of any information about their actual trustworthiness), whereas gullibility represents a failure to detect or act upon cues that indicate a person's untrustworthiness. However, it should be noted that even if trust and gullibility represent different personality characteristics, this does not necessarily mean that they represent different facial stereotypes. Additional studies are thus

needed to test whether people associate gullibility and trustfulness with different facial appearances or whether judgments of both traits are driven by the same facial features.

Finally, prior work on scamming suggests that the elderly are at higher risk of falling victim of scams (Burnes et al., 2017; Cohen, 2006). In contrast to these findings, we did not find that older adults were seen as more gullible than younger or middle-aged adults (Study 3). These diverging findings may be due to our broad conceptualization of gullibility. It may be the case that older adults are seen as more gullible in some situations (e.g., typical cases of fraud) but not others. Thus, additional work is needed to understand the relationship between age and (perceived) gullibility.

Conclusion

Across three studies, using both theory-driven and data-driven methods, we examined how people form gullibility impressions based on facial appearance. Results showed that (a) people show some agreement in what a gullible person looks like, (b) gullibility judgments are based on different facial features than judgments of trustworthiness and dominance, suggesting that they represent dissociable facial stereotypes, and (c) people rely on facial cues that are seen as indicators of positive intentions (e.g., smiles) and low levels of threat (e.g., babyfacedness) when forming gullibility impressions. These findings are in line with the view that gullible people are perceived as approachable and harmless.

References

Bonnefon, J. F., Hopfensitz, A., & De Neys, W. (2017). Can we detect cooperators by looking at their face? *Current Directions in Psychological Science*, *26*(3), 276–281. https://doi.org/10.1177/0963721417693352

Burnes, D., Henderson Jr, C. R., Sheppard, C., Zhao, R., Pillemer, K., & Lachs, M. S. (2017). Prevalence of financial fraud and scams among older adults in the United States: A systematic review and meta-analysis. *American Journal of Public Health*, *107*(8), E13–E21. https://doi.org/http://dx.doi.org/10.2105/AJPH.2017.303821

Button, M., Lewis, C., & Tapley, J. (2014). Not a victimless crime: The impact of fraud on individual victims and their families. *Security Journal*, *27*(1), 36–54. https://doi.org/10.1057/sj.2012.11

Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton: Princeton University Press.

Cohen, C. A. (2006). Consumer fraud and the elderly. *Journal of Gerontological Social Work*, *46*(3–4), 137–144. https://doi.org/10.1300/j083v46n03_08

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science*, *3*(5), 562–571. https://doi.org/10.1177/1948550611430272

Dotsch, R., Wigboldus, D. H. J., Langner, O., & Van Knippenberg, A. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychological Science*, *19*(10), 978–980. https://doi.org/10.1111/j.1467-9280.2008.02186.x

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES-A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods*, *42*(1), 351–362. https://doi.org/10.3758/BRM.42.1.351

Evans, A. M., & van de Calseyde, P. P. F. M. (2018). The reputational consequences of generalized trust. *Personality and Social Psychology Bulletin*, *44*(4), 492–507. https://doi.org/10.1177/0146167217742886

Gomulya, D., Wong, E. M., Ormiston, M. E., & Boeker, W. (2017). The role of facial appearance on CEO selection after firm misconduct. *Journal of Applied Psychology*, *102*(4), 617–635. https://doi.org/http://dx.doi.org/10.1037/apl0000172

Hehman, E., Sutherland, C. A. M., Flake, J. K., & Slepian, M. L. (2017). The unique contributions of perceiver and target characteristics in person perception. *Journal of Personality and Social Psychology*, *113*(4), 513–529. https://doi.org/10.1037/pspa0000090

Imhoff, R., Woelki, J., Hanke, S., & Dotsch, R. (2013). Warmth and competence in your face! Visual encoding of stereotype content. *Frontiers in Psychology*, *4*(June), 386. https://doi.org/10.3389/fpsyg.2013.00386

Jaeger, B., Evans, A. M., Stel, M., & van Beest, I. (2019). Explaining the persistent influence of facial cues in social decision-making. *Journal of Experimental Psychology: General*, *148*(6), 1008–1021. https://doi.org/10.1037/xge0000591

Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007). Social phishing. *Communications of the ACM*, *50*(10), 94–100. https://doi.org/10.1145/1290958.1290968

Krueger, J. I., Vogrincic-Haselbacher, C., & Evans, A. M. (2019). We need a credible theory of gullibility. In J. P. Forgas & R. F. Baumeister (Eds.), *Homo Credulus: The social psychology of gullibility*. New York: Taylor & Francis.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska directed emotional faces.* Psychology Section, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Martin, J., Rychlowska, M., Wood, A., & Niedenthal, P. (2017). Smiles as multipurpose social signals. *Trends in Cognitive Sciences*, *21*(11), 864–877. https://doi.org/10.1016/j.tics.2017.08.007

Mehu, M., Grammer, K., & Dunbar, R. I. M. (2007). Smiles when sharing. *Evolution and Human Behavior*, *28*(6), 415–422. https://doi.org/10.1016/j.evolhumbehav.2007.05.010

Modic, D., & Anderson, R. (2015). It's all over but the crying: The emotional and financial impact of internet fraud. *IEEE Security and Privacy*, *13*(5), 99–103. https://doi.org/10.1109/MSP.2015.107

Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Sciences*, *18*(11), 566–570. https://doi.org/10.1016/j.tics.2014.09.007

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Ormiston, M. E., Wong, E. M., & Haselhuhn, M. P. (2017). Facial-width-to-height ratio predicts perceptions of integrity in males. *Personality and Individual Differences*, *105*, 40–42. https://doi.org/10.1016/j.paid.2016.09.017

Rapoport, A., & Chammah, A. M. (1970). *Prisoner's Dilemma - A Study in Conflict and Cooperation*. The University of Michigan Press.

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, *35*, 651–665. https://doi.org/10.1111/j.1467-6494.1967.tb01454.x

Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion*, *9*(2), 260–264. https://doi.org/10.1037/a0014681

Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. F., & Downs, J. (2010). Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10*, 373–382. https://doi.org/10.1145/1753326.1753383

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428.

Sofer, C., Dotsch, R., Wigboldus, D. H. J., & Todorov, A. (2015). What is typical is good: the influence of face typicality on perceived trustworthiness. *Psychological Science*, *26*(1), 39–47. https://doi.org/10.1177/0956797614554955

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science*, *21*(3), 349–354. https://doi.org/10.1177/0956797610362647

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*(1), 105–118. https://doi.org/10.1016/j.cognition.2012.12.001

Sutherland, C. A. M., Rhodes, G., Burton, N., & Young, A. W. (2019). Do facial first impressions reflect a shared social reality? *British Journal of Psychology*. https://doi.org/10.1111/bjop.12390

Teunisse, A. K., Case, T. I., Fitness, J., & Sweller, N. (2019). I should have known better: Development of a self-report measure of gullibility. *Personality and Social Psychology Bulletin*, 014616721985864. https://doi.org/10.1177/0146167219858641

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*(1), 519–545. https://doi.org/10.1146/annurev-psych-113011-143831

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460. https://doi.org/10.1016/j.tics.2008.10.001

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science*, *26*(8), 1325–1331. https://doi.org/10.1177/0956797615590992

Witt, P. (2018). The top frauds of 2018. Retrieved January 19, 2020, from Federal Trade Commission website: https://www.consumer.ftc.gov/blog/2019/02/top-frauds-2018

Yamagishi, T., Kikuchi, M., & Kosugi, M. (1999). Trust, gullibility, and social intelligence. *Asian Journal of Social Psychology*, *2*(1), 145–161. https://doi.org/10.1111/1467-839X.00030